Watson, A. B. & Turano, K. A. (1994). The optimal motion stimulus. *Vision Research 1994* (in press).

The Optimal Motion Stimulus

Andrew B. Watson

NASA Ames Research Center,
Moffett Field, CA 94035-1000
beau@vision.arc.nasa.gov

Kathleen Turano

Wilmer Institute
The Johns Hopkins University School of Medicine
Baltimore, MD 21205
kathy@lions.med.jhu.edu

Correspondence:

Andrew B. Watson
MS 262-2
NASA Ames Research Center
Moffett Field, CA 94035-1000
beau@vision.arc.nasa.gov
(415) 604-5419

Date of this draft: October 20, 1994

Abstract

Contrast energy thresholds were measured for discriminating the direction of a drifting sinusoidal grating multiplied by an independently drifting space-time Gaussian (a generalized Gabor). We argue that the stimulus with the lowest contrast energy threshold identifies the receptive field of the most efficient linear motion filter. This optimal motion stimulus is found to be at 3 cycles/degree and 5 Hz, with a width and height of 0.44 degree and a duration of 0.133 seconds, corresponding to spatial and temporal bandwidths of 1.1 and 2.5 octaves, respectively. The spectral receptive field is aligned more nearly to the Cartesian axes than to the velocity contour.

Theory

Early thinking on the nature of motion detecting mechanisms in human vision was dominated by the notion of a matching process operating over space and time. Activity at one point in space and time was matched to activity at another point in space and time, and motion between the two points was thereby inferred. The matching was typically accomplished by conveying the activity from the first point to the second, with a delay corresponding to the putative speed of motion, and multiplying the two activities (Barlow & Levick, 1965; Reichardt, 1961; Reichardt, 1986). Advances in visual physiology (Hamilton, Albrecht & Geisler, 1989), psychophysics (Adelson & Movshon, 1982), and mathematical analysis of the motion problem (Crick, Marr & Poggio, 1981; Watson & Ahumada, 1983; Watson, Ahumada & Farrell, 1986) have led to a new model, in which motion signals are first extracted by means of linear filters (Watson & Ahumada, 1983). This motion filter model has been used to compute local image velocity (Watson & Ahumada, 1985), " opponent" motion signals (van Santen & Sperling, 1985) , "motion energy" (Adelson & Bergen, 1985) and "motion magnitude" (Watson, 1990), and to detect gradients in the motion field (Watson & Eckert, 1994).

The characteristic feature of the motion filter, when viewed in the three-dimensional spatiotemporal frequency domain, is that its passband or "spectral receptive field" lies predominantly in one half of the positive temporal frequency half-volume (Watson & Ahumada, 1983). Viewed in one temporal and one spatial dimension (with the spatial dimension aligned with the preferred direction of motion), the motion filter passband lies predominantly in two opposing quadrants of the frequency domain (Fig. 1). This spectrum corresponds to a receptive field (or impulse response) that appears oriented in space-time. Beyond this fundamental structural feature, there are many detailed questions that may be asked about the motion filters in human vision. What are their spatial and temporal bandwidths, or corresponding height, width, and duration of the receptive field? What is the detailed shape and orientation of the spectrum or receptive field?

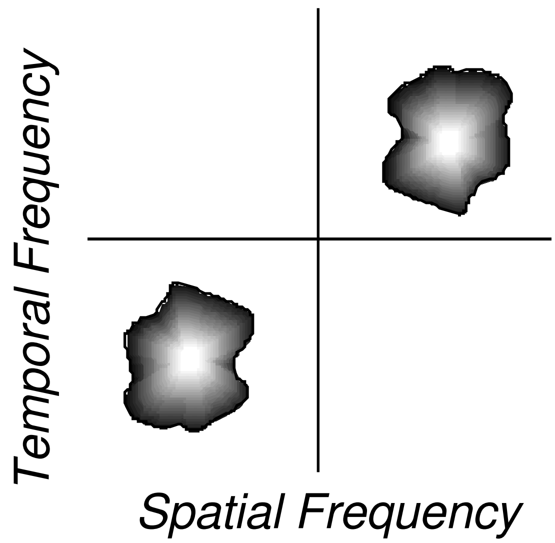

Figure 1. Frequency spectrum of generic motion filter.

Using both spatial summation (Anderson & Burr, 1985; Anderson & Burr, 1987; Anderson & Burr, 1991) and masking experiments (Anderson & Burr, 1989; Anderson, Burr & Morrone, 1991; Burr, Ross & Morrone, 1986), Anderson, Burr and Morrone have provided considerable information on the shape of the motion filters. By examining variations in sensitivity as a function of length and width of a moving test grating, and as functions of spatial frequency and orientation of drifting and randomly phase-changing masks, they have derived estimates of some of the dimensions of the motion receptive field. Their masking results suggested spectral receptive fields that are quite broad in temporal frequency and moderately broad in spatial frequency. Summation data indicate spatial receptive fields that are roughly as tall as they are wide (an aspect ratio of 1), and a width (defined as two standard deviations of a Gaussian window) that increases from about 0.1 cycle at 0.1 cycle/degree to 0.5 cycle at 10 cycles/degree. These widths result in rather broad bandwidths. In octave terms, their narrowest bandwidth, at 10 cycles/degree, is 2.6 octaves. The octave bandwidth for the two lower spatial frequencies cannot be computed because the lower half-amplitude point is at a negative frequency. These bandwidths are substantially larger than the median for V1 neurons of 1.4 octaves, though physiological bandwidths are highly variable (De Valois, Albrecht & Thorell, 1982). The masking data suggest widths about twice as great, and thus bandwidths that are considerably narrower, but this comparison is complicated by the fact that rather different receptive field models were used to analyze summation and masking data.

The temporal dimension has been examined by means of masking experiments, which yielded very broad bandwidths (Anderson & Burr, 1985). Masking functions did not peak at the test frequency and showed only weak evidence for more than a single temporal mechanism.

These results clarify considerably our picture of the spectral receptive field. But one objection to many of these experimental approaches is that they used only a single spatial frequency at a range of temporal frequencies, or a single temporal frequency at a range of spatial frequencies, and that they therefore assume a spectral receptive field that is positive-separable[1] in spatial and temporal frequency. (an exception is Burr, Ross, and Morrone, 1986). To illustrate this point, Fig. 2 shows three possible spectral receptive fields, all three of which have the same spatial and temporal frequency bandwidths. One of the three (*a*) is separable in spatial and temporal frequency, and is therefore oriented along the Cartesian axes. Another (*b*) is oriented along the line of constant velocity, and might therefore be described as "velocity tuned." The third (*c*), oriented orthogonal to the velocity contour, has no simple interpretation but is nonetheless a logical and physical possibility (see (Fleet & Langley, 1993) for a possible interpretation).

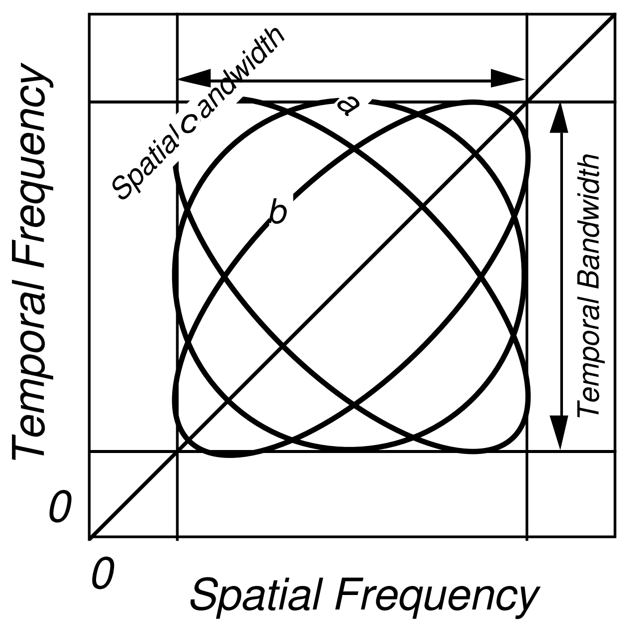

Figure 2. Three possible passbands with identical spatial and temporal bandwidths.

To address this and other gaps in our knowledge of the receptive field of the motion filter, we have adapted a technique developed earlier to estimate the shape of the receptive fields involved in luminance contrast detection (Watson, Barlow & Robson, 1983). In that study of "what does the eye see best," it was argued that for a fixed linear receptive field, the most efficiently detected stimulus is one that matches the shape of the receptive field. Efficiency is measured as the inverse of the threshold contrast

[1]Motion filters are by definition not separable in space and time, or spatial and temporal frequency(Watson & Ahumada, 1983). They may, however, be separable when only positive temporal frequencies are considered. We call this "positive-separable".

energy. Contrast energy is the integral of the square of the contrast waveform. The experimental approach, then, is to survey a wide range of plausible stimuli to discover which is detected with least contrast energy. The waveform of this optimal stimulus putatively identifies the shape of the receptive field. In practice, because all stimuli cannot be investigated, the search is confined to some plausible parameterized family of candidates.

We modify this approach in only one respect. Because we are interested in the shape of the *motion* receptive field, the thresholds we measure are for a direction-discrimination judgment. On each trial, the stimulus moves either right or left, and the observer must try to discriminate this direction.

This optimization approach relies on two observations. The first, which is a mathematical truth, is that if there is a linear motion filter, its receptive field will correspond to the optimal stimulus. This result is a direct inversion of the familiar matched filter theorem, which states that the ideal detector of a signal known exactly is a filter whose impulse response matches the signal (Duda & Hart, 1973; Green & Swets, 1966; Watson, et al., 1983). The second observation is that, since the linear filter is ideal, it is a likely candidate for a motion sensor, particularly at the early stages of vision. This expectation is bolstered by extensive evidence for cortical neurons that act to a good first approximation as linear motion filters. But we must acknowledge at the outset that in human vision 1) linear motion filters may not exist, and 2) even if they do exist and are well characterised by our procedure, that other, less efficient non-linear motion sensors may exist.

To select a plausible search space, we take note of the filter model cited earlier, which often employs a Gabor function in the space domain, and the results cited above which indicate a receptive field that is local in both two-dimensional space and frequency. As discussed below, this leads to a stimulus family that we call "generalized Gabors."

Stimuli

The family of stimuli that we employ can be described either in their space-time or frequency domain aspects. In space-time, our stimulus consists of a drifting sinusoidal grating, with a frequency of $\mathbf{f} = [f_x, f_y, f_t]$ (and thus a velocity of $f_t/[f_x, f_y]$ windowed by a Gaussian aperture with spatial and temporal scales of s_x, s_y, and s_t. The Gaussian aperture may itself move with a velocity $[a_x, a_y]$. We will call these stimuli "generalized Gabors." From their context in the theory of modulation, we will refer to the grating as the *carrier* and the Gaussian as the *aperture*. Fig. 3 provides some $[x,t]$ images of possible generalized Gabor stimuli. In the upper two images, the grating moves to the right at 1 deg/sec and the aperture is stationary. The two panels differ only in horizontal and vertical scales. In the lower two panels, the aperture either moves with the same (C) or opposite velocity (D) as the grating. The latter two examples address one question of particular interest: does the aperture move with the carrier in the human motion receptive field, or is it stationary? As we shall see, this is equivalent to asking whether the spectral receptive field is aligned with the Cartesian axes, and thus possibly positive-separable.

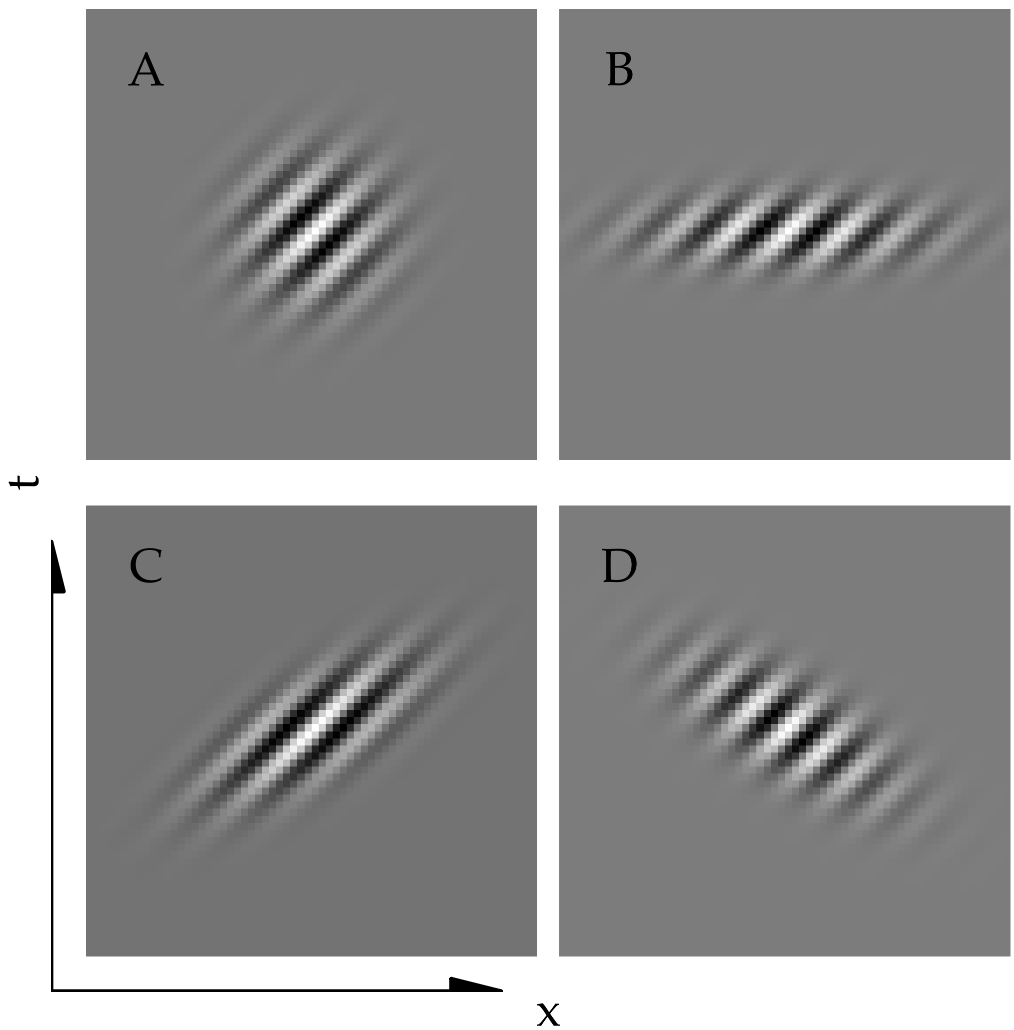

Figure 3. Examples of generalized Gabor stimuli. The spatial and temporal extents are one degree and one second. Unless noted, parameters are: $f_x = 8$ cycles/deg, $f_t/f_x = 1$ deg/sec, $s_x = 0.25$ deg, $s_t = 0.25$ sec, $a_x = 0$ deg/sec. Varying parameters are: B) $s_x = 0.5$ deg, $s_t = 0.125$ sec, C) $a_x = 1$ deg/sec, D) $a_x = -1$ deg/sec.

The three-dimensional Fourier transform of the generalized Gabor can be easily derived in the following way. The transform of the carrier grating is simply a pair of impulses at $\pm \mathbf{f}$. The transform of the 3D Gaussian aperture is itself a 3D Gaussian. Multiplication of the carrier and aperture corresponds to convolution of their Fourier transforms, and convolution of a Gaussian with an impulse corresponds to placing a copy of the Gaussian at the location of the impulse. The result is therefore a pair of 3D Gaussians located at $\pm \mathbf{f}$. Finally, changes in the width, height, duration, and velocity of the aperture correspond to magnifications and shears of the 3D Gaussian, which correspond to complementary magnifications and shears in the frequency domain.

To be specific, consider a "unit" 3D Gaussian in space-time, with a scale of 1 in each dimension, which we write as

$$\exp(-\pi \mathbf{x}'\mathbf{x}) \tag{1}$$

where $\mathbf{x} = [x, y, t]$, and where the prime symbol indicates matrix transposition. For this unit 3D Gaussian, a surface of constant value of $\exp(-\pi)$ is a sphere of radius 1. We shall say that its width, height, and duration are all 1. Changing the scales of the Gaussian, and putting it in motion, can be represented by linear geometric transformations of space-time. Scaling is described by a matrix

$$\mathbf{S} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_t \end{bmatrix} . \tag{2}$$

where s_x, s_y, and s_t describe the new width, height, and duration.

While some analyses of motion sensing have made an analogy between orientation in space and velocity in space-time(Adelson & Bergen, 1985; Burr, et al., 1986), this is not strictly correct. Motion corresponds to a *shearing* transformation of space-time rather than a rotation. To see this, consider just two dimensions (x and t) and imagine a stationary signal $f(x,t)$. If this signal is placed in motion at speed r, it may be written as $f(x-rt,t)$. This corresponds to a transformation of the coordinate vector $[x,t]'$ to $\mathbf{M}\,[x,t]'$ where

$$\mathbf{M} = \begin{bmatrix} 1 & -r \\ 0 & 1 \end{bmatrix} \tag{3}$$

This is a shearing transformation, rather than a rotation. In three dimensions, with horizontal and vertical speeds r_x and r_y, the motion shear matrix is

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & r_x \\ 0 & 1 & r_y \\ 0 & 0 & 1 \end{bmatrix} . \tag{4}$$

When motion precedes scaling, the complete transformation $\mathbf{T}$ is the product of motion and scaling transformations $\mathbf{M}$ and $\mathbf{S}$,

$$\mathbf{T} = \mathbf{M}\mathbf{S} = \begin{bmatrix} s_x & 0 & r_x s_t \\ 0 & s_y & r_y s_t \\ 0 & 0 & s_t \end{bmatrix} \tag{5}$$

After transformation by the matrix $\mathbf{T}$, the unit Gaussian may be written

$$\exp\left(-\pi\, \mathbf{x}' \mathbf{C}^{-1} \mathbf{x}\right) \tag{6}$$

where

$$\mathbf{C} = \mathbf{T}\mathbf{T}' \tag{7}$$

Transformation of space-time by the matrix $\mathbf{T}$ corresponds to a transformation of the frequency domain by the matrix $(\mathbf{T}')^{-1}$ and the corresponding Gaussian in the frequency domain is

$$|\mathbf{T}| \exp\left(-\pi\, \mathbf{u}' \mathbf{C}' \mathbf{u}\right) \tag{8}$$

where $\mathbf{u} = [u,v,w]$ is the 3D frequency coordinate.

This general formula includes the simple cases in which an expansion in space-time results in a contraction by an equal factor in the frequency domain and in which a rotation in space-time results in an equal rotation in frequency. This is illustrated in Fig. 4 in which we picture ellipses corresponding to a particular set of scales and speeds, as well as the corresponding ellipses in the frequency domain. All ellipses represent $\exp(-\pi)$ contours of the corresponding Gaussians. For simplicity, we show only two dimensions. Note that the space-time and frequency ellipses are always orthogonal. From the left panel to the right, the duration is shorter and the speed is greater. This yields a frequency ellipse that is broader in temporal frequency and more steeply inclined[2].

---

[2]Since the contours in Fig. 4 were produced by linear transformations of a circle, they must all be ellipses. Thus, even though motion is represented by a shear, for the special case of a Gaussian this is equivalent to a particular magnification and rotation. It sometimes proves convenient to know what this magnification and rotation are, so we present them here for reference. In general, given a linear transformation $\mathbf{T}$, a circle is transformed into an ellipsoid with principal axes equal to the eigenvectors of $\mathbf{C} = \mathbf{T}\mathbf{T}'$, with lengths equal to the square roots of the eigenvalues. This transformation is equivalent to a magnification by the diagonal matrix of lengths, followed by a rotation to the direction of the first eigenvector.

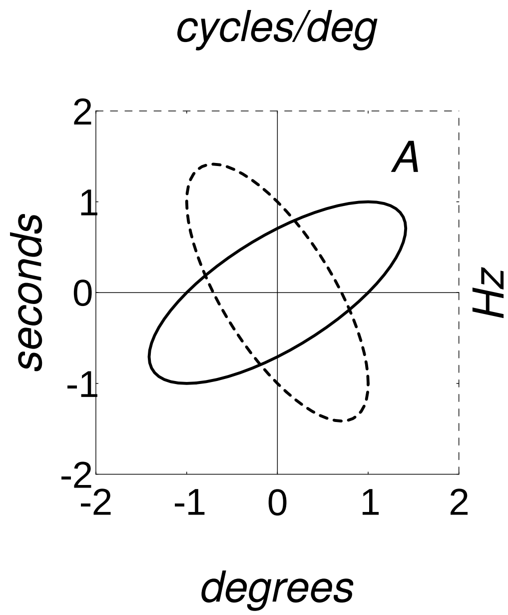

Figure 4. Space-time and frequency domain ellipsoids corresponding to particular aperture scales and motions. The solid ellipse is the constant-value (exp(-π)) contour of the space-time Gaussian; the dashed ellipse is the contour for the corresponding frequency Gaussian. A) $s_x = 1$ deg, $s_t = 1$ second, $a_x = 1$ degree/second. B) $s_x = 1$ degree, $s_t = 0.5$ second, $a_x = 2$ degree/second.

Note that the Gaussian and its modulation may be considered separately and that the stimulus bandshape, as distinct from its location, is determined entirely by the 3D Gaussian. This family of generalized Gabors is attractive in part because it can be easily expanded or contracted in each of the three dimensions and may also have its major axis oriented along an arbitrary direction. As noted above, certain directions have theoretical interpretations of particular interest. The translations, of course, serve to center the ellipsoid on a particular three dimensional spatiotemporal frequency.

Finally, using the notation developed above, we can write the luminance distribution for our generalized Gabor stimulus as:

$$\mathrm{L}(\mathbf{x}) = \overline{L}\left[1 + \mathrm{c}(\mathbf{x})\right] \tag{9a}$$

$$\mathrm{c}(\mathbf{x}) = m\ \exp\left[-\pi\,\mathbf{x}'\mathbf{C}^{-1}\mathbf{x}\right]\ \cos\left[2\pi\,\mathbf{f}'\,\mathbf{x}\right] \tag{9b}$$

where $\overline{L}$ is the mean luminance, $\mathrm{c}(\mathbf{x})$ is the contrast waveform, and m is the peak contrast.

Contrast energy

The contrast energies of our transformed Gaussian stimuli are easily computed. We first note that, by Parseval's Theorem, the energy of a signal is equal to the energy of its Fourier transform. The transforms of our stimuli are in every case a pair of transformed Gaussians, displaced to the two loci of the 3D sinusoid. Clearly the energy in the pair of Gaussians does not depend upon their locations (provided they do not overlap), and hence the energy does not depend upon the spatiotemporal frequency of our stimulus, only upon the aperture.

Next we note that the energy of a unit Gaussian of scale a in one dimension is $a/\sqrt{2}$. Each of our transformed Gaussians, we have seen, is a unit Gaussian subjected to scaling and shearing. The shearing does not affect the energy (again, assuming no overlap) so we can ignore it. The total energy is then the product of the three energies of the three separable Gaussians, with scales s_x, s_y and s_t, times two to account for the two Gaussians, and multiplied by $m^2/4$ because the amplitude of each 3D Gaussian, before squaring, is $m/2$.

$$E = 2^{-5/2} m^2 \left(s_x s_y s_t \right) \quad . \tag{10}$$

Note that this quantity depends only on the spatial and temporal scales, and not on the velocity of the grating or the aperture, or upon the carrier frequency.

Methods

Stimuli were computed in advance as digital movies with eight bit precision. Movie resolution was 256 by 256 pixels by 16 frames. Each movie was stored in the framebuffer memory of a PIXAR II Image Computer, and could be presented at a selected frame rate at a selected contrast. Contrast control and display linearization were accomplished by means of look-up-tables just prior to the 10 bit digital-to-analog converters of the framebuffer controller. In these experiments, frame rate was always 30 Hz. This value was chosen as the best compromise between excessive storage and computation requirements for each movie and the potential for aliasing. Consideration of the spatial and temporal parameters of our stimuli shows that none were significantly aliased at this frame rate. Display mean luminance was 40 cd/m^2. Stimuli were presented on a dark background in an otherwise darkened room. Viewing was monocular with the dominant eye from a distance of 48.4 cm, yielding an image size of 8x8 deg. The non-dominant eye viewed the display through a diffuser. Three observers (one naive) took part. All observers wore their normal spectacle correction.

Data were collected with a two-alternative forced-choice QUEST staircase procedure (Watson & Pelli, 1983), and thresholds were subsequently estimated as the 82% correct point of a fitted Weibull function (Watson, 1979). On each trial, the stimulus moved randomly either left or right, and the observer attempted to identify this direction. When both grating and aperture drifted, the observer identified the grating direction. Feedback was provided.

In these experiments the carrier frequency was always horizontal (f_y=0), and carrier speed was of necessity horizontal. Our search strategy was to optimize the remaining parameters in the following order: spatial frequency (f_x), carrier speed (f_t/f_x), duration (s_t), width and height (s_x and s_y), and aperture speed (a_x and a_y).

Results

Spatial Frequency

For the first series of measurements, which looked for the optimal spatial frequency, it was necessary to make initial guesses for the values of the other parameters. The horizontal and vertical scales (s_x and s_y) were both set equal to 2.66 cycles of the carrier, and the temporal scale (s_t) was set to 4 frames (0.133 sec). The speeds were set to result in a constant temporal frequency of 4 Hz. These numbers were based loosely upon the optima obtained by Watson *et al.* (1983).

Figure 5 shows the results for three observers. Results are plotted as -$\log_{10}$ of contrast energy, which is proportional to the $\log_{10}$ of efficiency under the assumption of a flat noise spectrum. Efficiency declines markedly below 2 and above 4 cycles/degree, and between these points there is a rather flat optimum. Frequencies of 2, 3, and 4 are equally efficient within measurement error, so we selected 3 cycles/degree as the optimum from which the search would continue in another dimension.

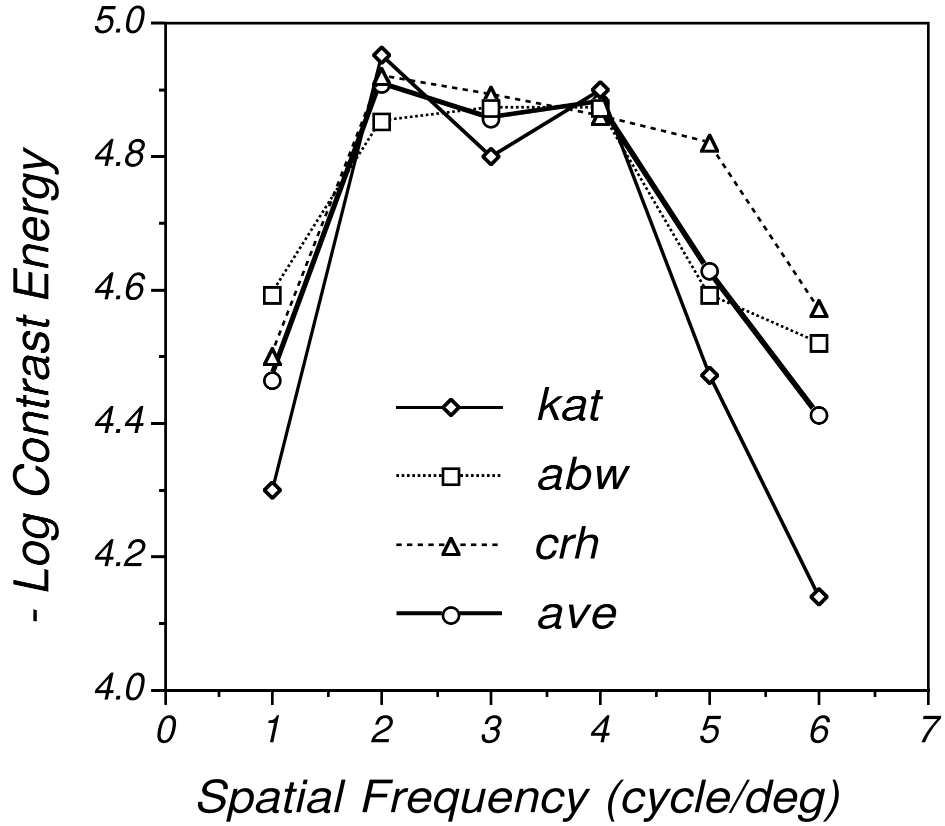

Figure 5. Contrast energy thresholds for various spatial frequencies.

The peak at approximately 3 cycles/deg differs from the value of around 8 cycles/deg previously obtained for efficiency of simple detection (Watson, et al., 1983) . This difference is consistent with the common observation that the motion pathway is preferentially sensitive to low spatial frequencies (Kulikowski & Tolhurst, 1973; Tolhurst, 1973; Tolhurst, 1975; Watson & Robson, 1981; Watson, Thompson, Murphy & Nachmias, 1980). Consistent with this view, Watson *et al.* (Watson, et al., 1980) showed that direction discrimination thresholds are somewhat below detection thresholds at around 5 Hz.

Carrier Speed

With spatial frequency fixed at 3 cycles/degree, and all other parameters fixed at their initial values (see above), we varied the carrier speed (f_t/f_x.). Results are shown for three observers in Fig. 6.

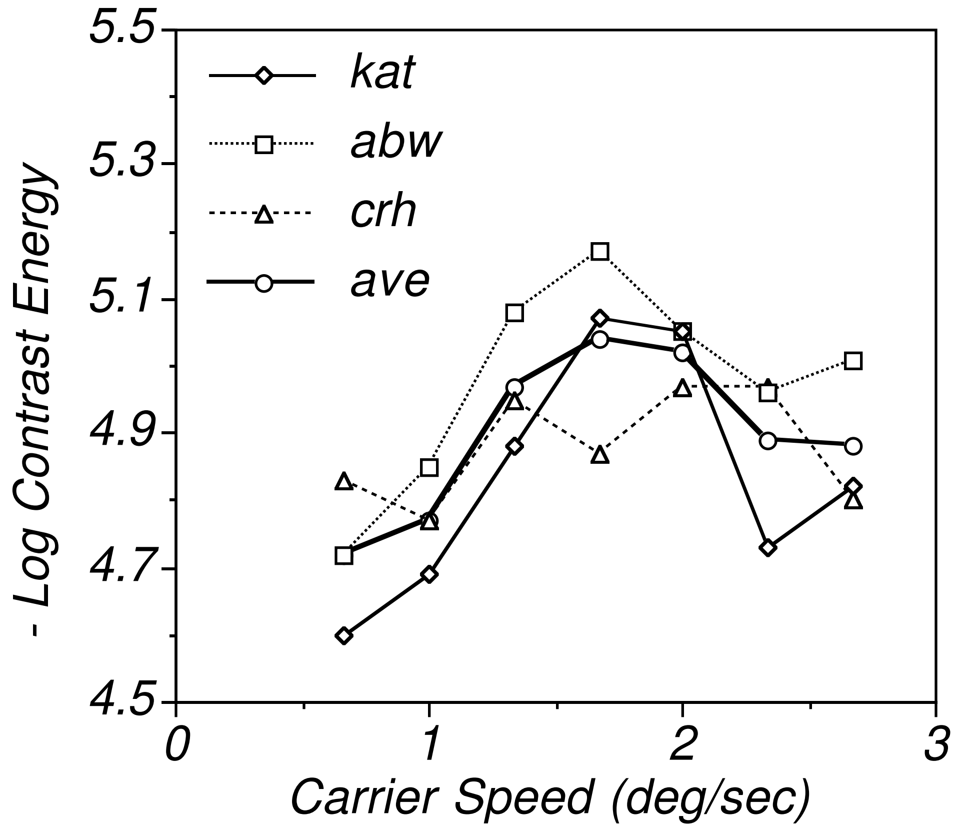

Figure 6. Contrast energy thresholds for various carrier speeds.

Considering the average of the three observers, the optimum occurs at a speed of 1.67 deg/sec. Values of 1.33 and 2.0 deg/sec are detected with an efficiency that is not significantly different. The optimum corresponds to a temporal frequency of 5 Hz, essentially the same as the value of 4 Hz determined by Watson *et al.* (1983).

Duration

With spatial frequency and carrier speed fixed at their optimal values, we next varied duration (s_t). Results are shown in Fig. 7.

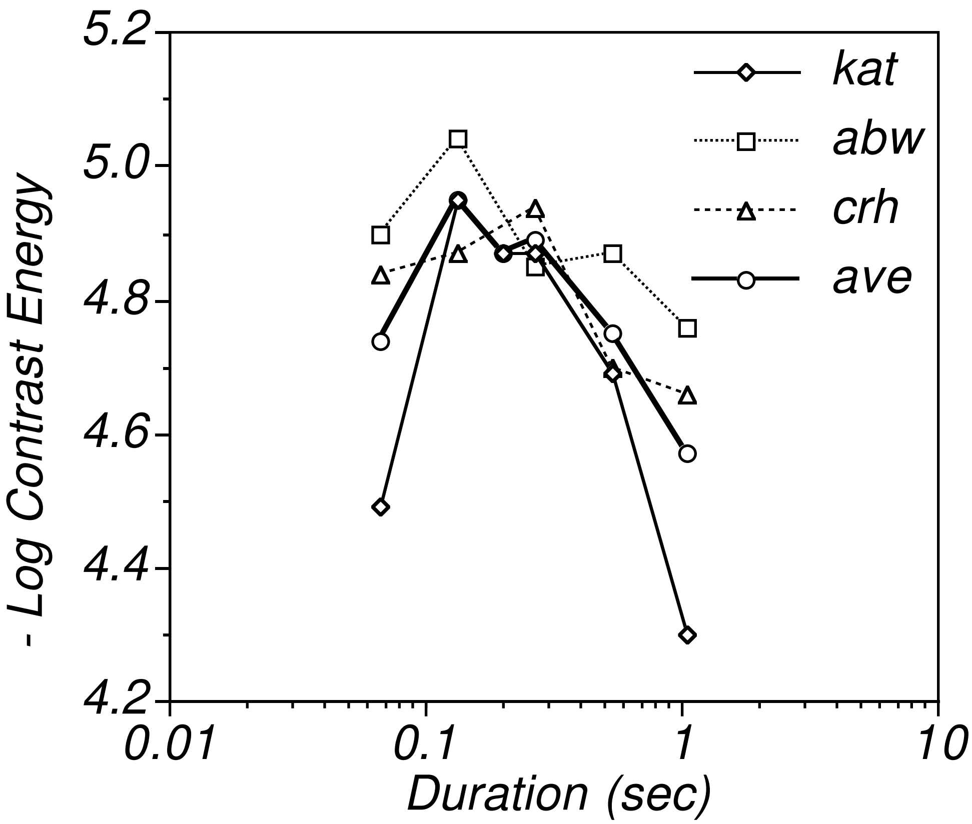

Figure 7. Contrast energy threshold as a function of duration (s_t).

There is some variability among observers, particularly at the longest and shortest durations, but the average peaks at about 0.133 sec. This rather brief duration corresponds to a broad temporal frequency spectrum (a scale of 7.5 Hz, or a half-amplitude, full bandwidth of 2.5 octaves). This is in rough agreement with estimates derived by Anderson and Burr (1985) from temporal masking studies.

Width and Height

In one set of measurements, pictured in Fig. 8, we simultaneously varied width and height of the aperture (s_x and s_y) while all other parameters remained at their current optima. The results show a very clear decline at larger sizes, and a more modest decline at the smallest sizes. The optimum of the average of the two observers is at 0.44 deg. This corresponds to a spatial frequency bandwidth of 1.1 octaves.

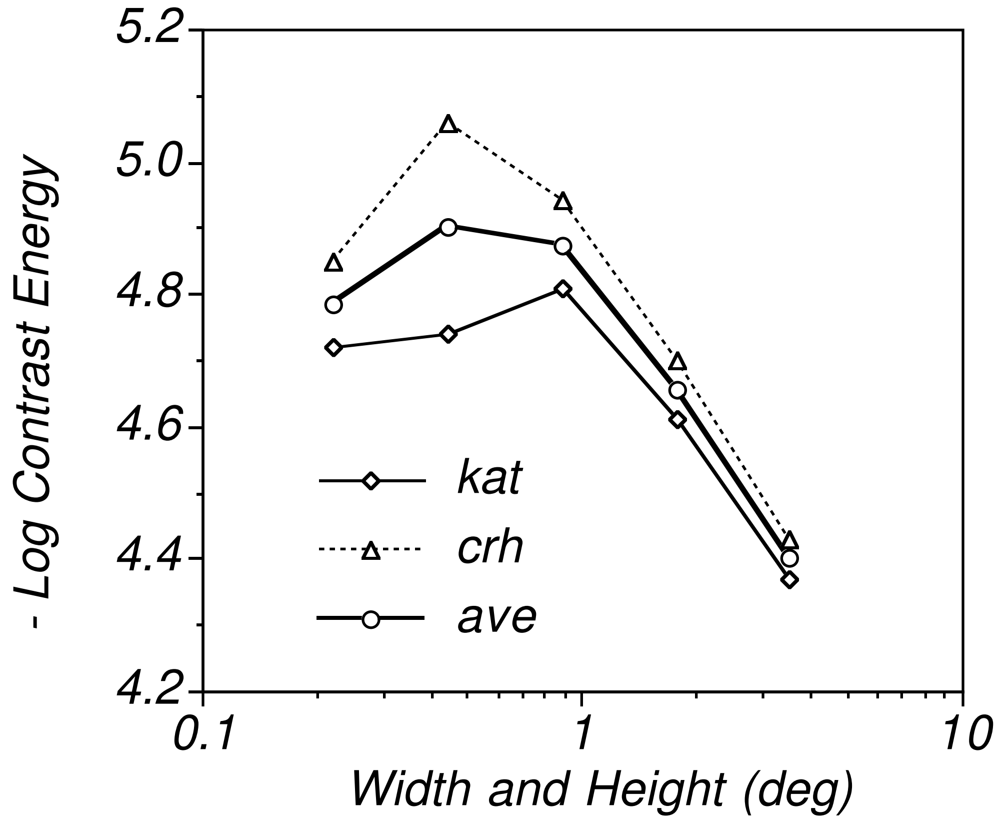

Figure 8. Contrast energy thresholds as a function of the width and height of a circularly symmetric Gaussian aperture.

In additional measurements, we varied either width or height, while the other dimension was fixed at the optimum of 0.44 deg. These variations produced less than 0.1 log unit change in threshold contrast energy. This illustrates the complicating effects of probability summation over space (see discussion below) and shows that our estimates of receptive field size are only approximate. The approximate equivalence of width and height optima also agrees with previous masking and summation data (Anderson & Burr, 1991; Anderson, et al., 1991).

Aperture Speed

In our final experiment we varied the velocity of the aperture, while the other parameters of the stimulus were fixed at their optimal values. We used primarily horizontal aperture motion (the same axis as the grating motion), but in one case examined upward motion.

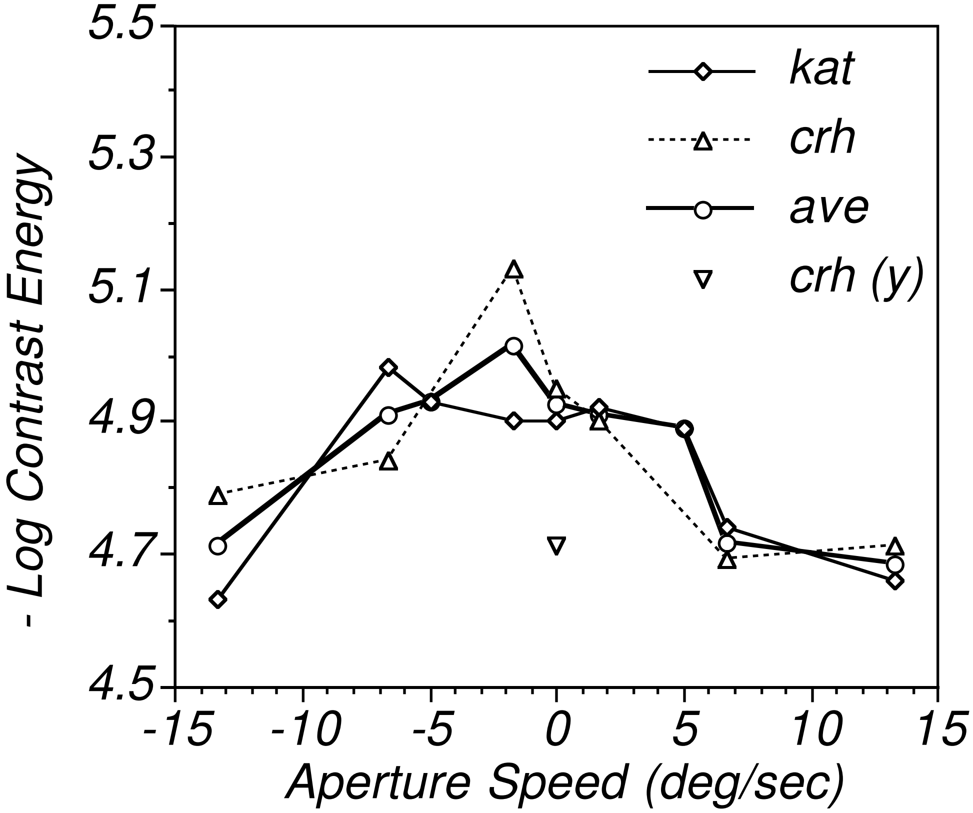

Fig. 9 Contrast energy threshold as a function of the speed of the Gaussian aperture. All points are for horizontal motion (a_x) except for the single point labeled "crh (y)" which is for upward motion (a_y) at 13.3 deg/sec.

Fig. 9 shows a broad optimum extending from around -6 to 6 deg/sec. The optimal velocity is approximately zero but could be either equal (1.67 deg/sec) *or opposite* (-1.67) to the grating speed. Consideration of the spectra corresponding to these three conditions may be edifying. As shown in Fig. 10, and as discussed earlier, motion in space-time results in a shear (not a rotation) of space-time and a related shear in frequency. In graphical terms, this means that variation in the aperture speed will produce slight changes in the orientation of the spectrum, but will not rotate it to the orientation of the velocity line. It should be clear that the degree of possible rotation is determined by the aspect ratio in space and time: rotations are most easily accomplished when the spectrum is narrow in temporal frequency, and broad in spatial frequency. This is effectively the opposite of what holds for the optimal spectrum. Another perspective on this limitation is that variations of aperture speed do not alter the spatial frequency spectrum. The bandwidth of this spatial spectrum, which is observed to be rather narrow, constrains the rotations that can be achieved. To summarize, for the optimal signal, spatial bandwidth is too narrow and temporal bandwidth too broad to produce a spectral receptive field that is tuned for "velocity."

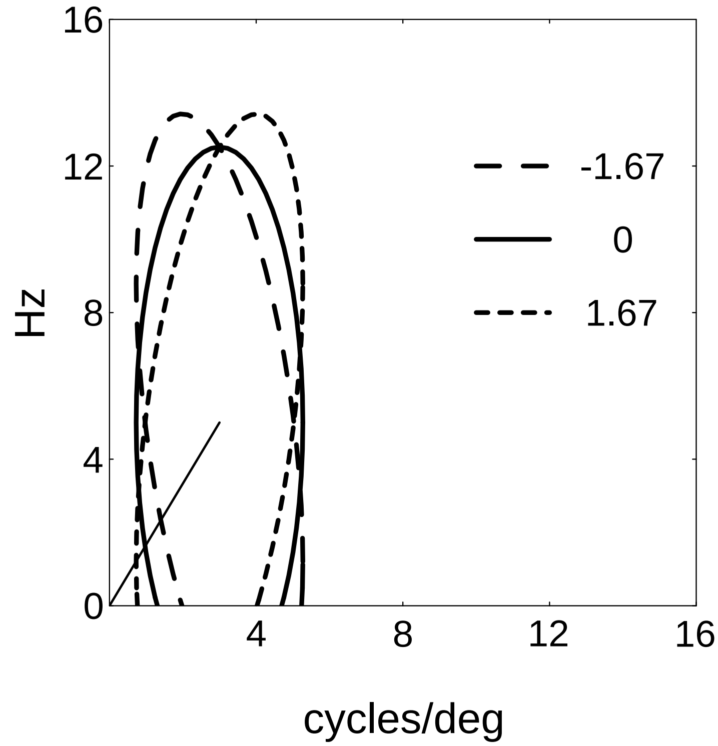

Figure 10. Frequency spectra corresponding to aperture motion of -1.67, 0, and 1.67 deg/sec. The oblique line from the origin corresponds to the carrier speed of 1.67 deg/sec, and extends to the carrier frequencies of 3 cycles/deg and 5 Hz. The aperture scales are 0.44 deg and 0.133 sec. Lines are isoamplitude contours at exp($-\pi$).

The optimal motion stimulus

We summarize the outcome of our sequential optimization of the stimulus parameters in Table 1. We have labeled as approximate those parameters which exhibited a broad optimum, or those that were not studied extensively (such as a_y).

Parameter	Value	Unit	Notes
fx	3	cycle/deg	
fy	0	cycle/deg	fixed
ft	5	Hz	1.67 deg/sec
sx	0.44	deg	approximate
sy	0.44	deg	approximate
st	0.133	sec	
ax	0	deg/sec	approximate
ay	0	deg/sec	approximate

Table 1. Parameters of optimal motion stimulus.

This optimal stimulus is rendered as an x-t image in Fig. 11, and as a 3D spectrum in Fig. 12.

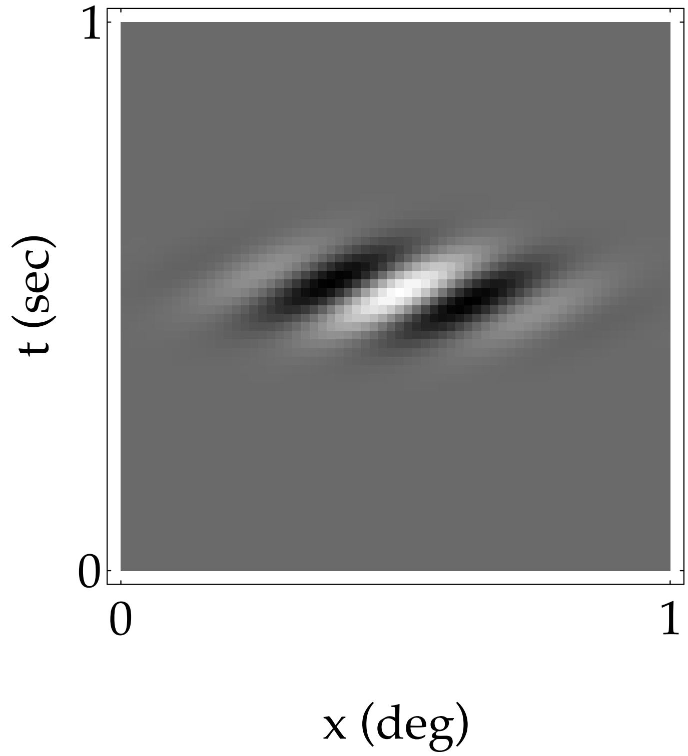

Figure 11. Space-time image of the optimal motion stimulus.

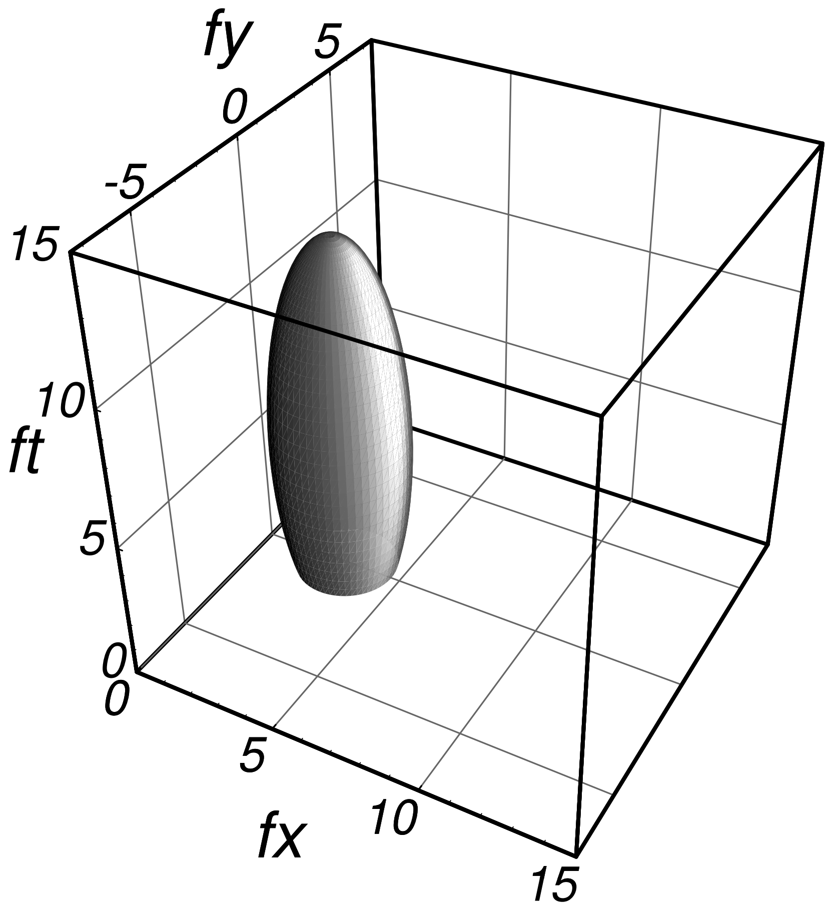

Figure 12. Three-dimensional amplitude spectrum of the optimal motion stimulus.

Discussion

Spectral Requirements for a "velocity-tuned" Sensor

We noted previously that with the spatial and temporal bandwidths we observed, it is impossible to produce a sensor whose spectral receptive field is "velocity-tuned", that is, aligned with the velocity contour. Here we generalize this observation somewhat. Consider a spectral receptive field following the generalized Gabor model, centered at (f_x,f_t) and with linear spatial and temporal bandwidths b_x and b_t (Fig. 13). To be aligned with the velocity contour (diagonal line in Fig. 13), the bandwidths must be in the same ratio as the frequencies:

$$\frac{b_t}{b_x} = \frac{f_t}{f_x} \quad . \tag{11}$$

Rearranging terms, we see that the ratio of bandwidth to center frequency must be equal for both spatial and temporal domains

$$\frac{b_t}{f_t} = \frac{b_x}{f_x} \quad . \qquad (12)$$

Another way of saying this is that spatial and temporal log bandwidths must be equal. This requirement is violated by typical psychophysical and physiological measurements which indicate spatial log bandwidths much narrower than temporal (an exception are the estimates of Anderson and Burr (1985) at low spatial frequencies). Finally we should note that since our methods reveal only the most efficient detector, it is possible that less efficient velocity-tuned detectors exist.

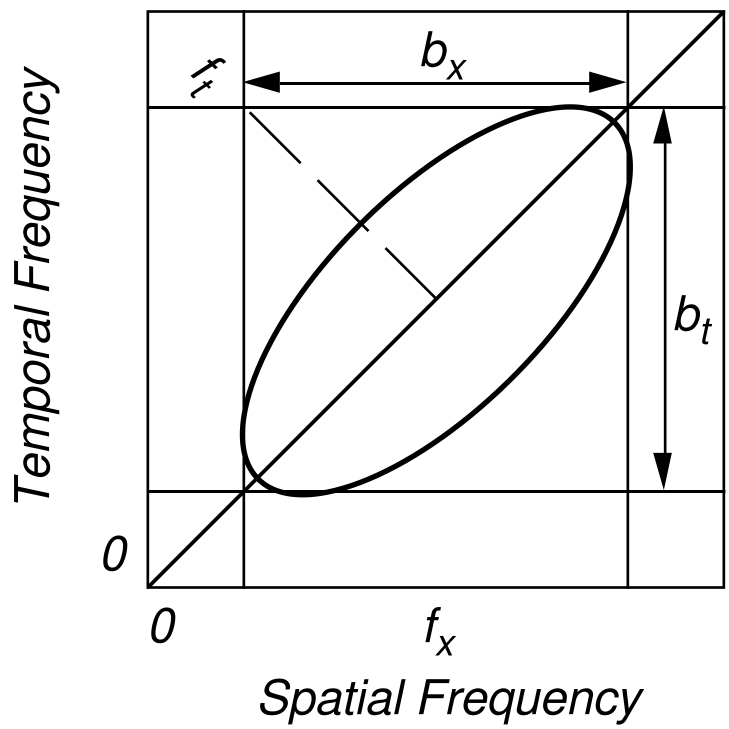

Figure 13. Bandwidth constraints for a spectral receptive field aligned with the velocity contour.

Comparison with results of Anderson and Burr (1991)

It is of interest to compare our results with those of Anderson and Burr (1991), who examined the effect on direction discrimination thresholds of the height and width of a Gaussian-windowed drifting grating. In most respects our stimuli and experimental methods closely resemble theirs, although our selection of stimulus parameters and data analyses are different.

Anderson and Burr collected thresholds for both detection and direction discrimination. We have extracted all of the discrimination data from their Figs. 1-4 by scanning, digitally measuring, and appropriately scaling the figure images. As a test of the accuracy of our data extraction methods, we have computed the standard deviation of our estimates of the x-coordinates from the twelve graphs (all twelve share the uppermost 11 x-coordinates). This value, averaged over 11 coordinates, was .005 $\log_{10}$ units. The contrast thresholds were converted to contrast energy thresholds by eqn. 10. Results are plotted in Fig. 14.

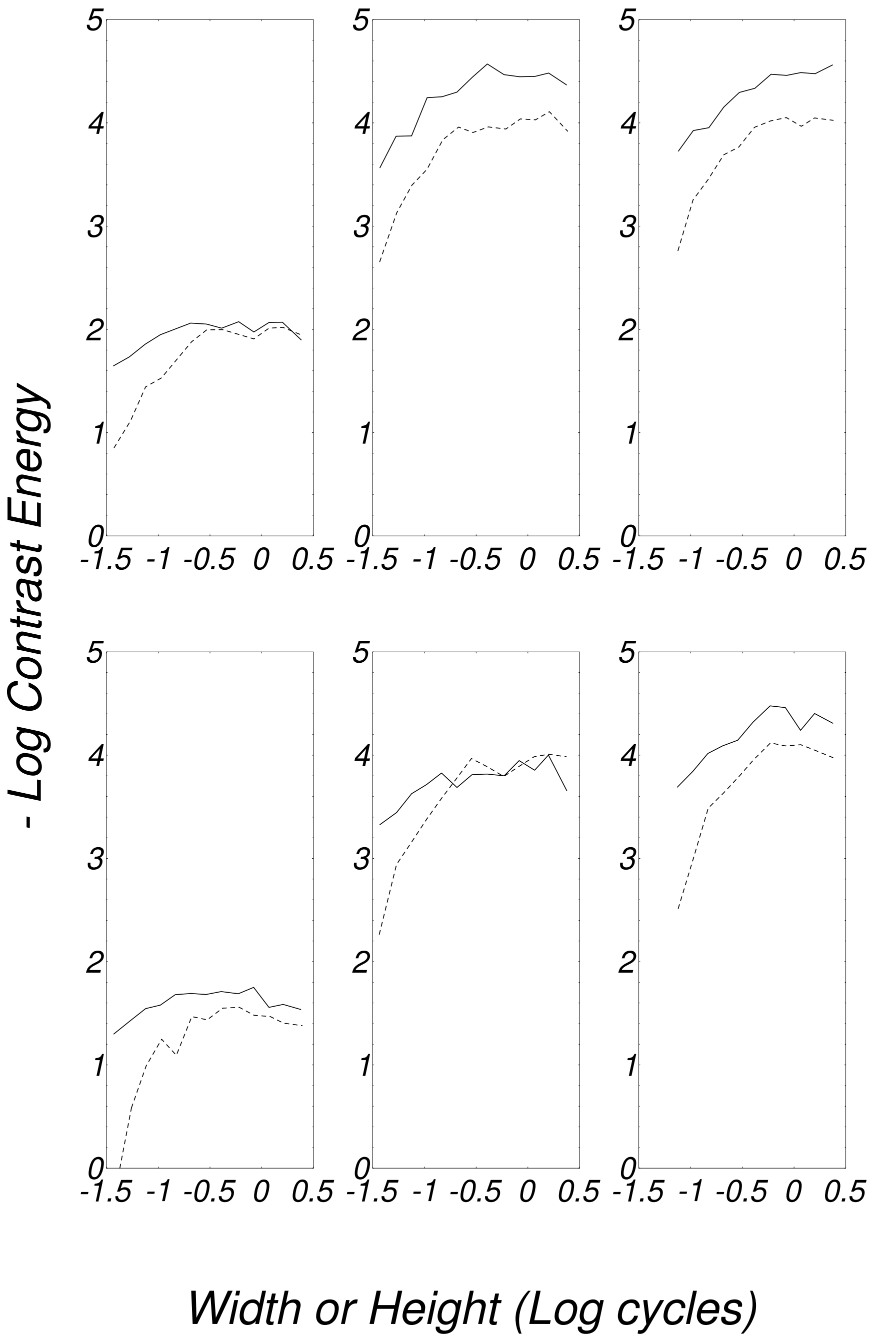

Figure 14. Contrast energy thresholds for direction discrimination derived from the data of Anderson and Burr (1991). The three upper panels are for observer SA, the lower three for AP. Solid lines show results for varying height, dashed lines for varying width. From left to right the spatial frequencies were 0.1, 1, and 10 cycles/degree.

The peak values attained in their data are around 4.6 log units. This is about 0.5 log unit below our best values. Several factors may contribute to this discrepancy. First, their frequencies of 1 and 10 cycles/degree lie on either side of our optimum, and from our data we expect as much as a 0.4 log unit decline from this effect alone (see Fig. 3). Second, they used a duration (s_t) of 0.827 sec, rather far from our optimum of 0.133 (Fig. 7), and we expect a further decline of perhaps 0.25 log units from this source. Third, they used a mean luminance of 400 cd/m^2, one log unit above ours. Since we are somewhere between DeVries-Rose and Weber regimes (van Nes & Bouman, 1967), we expect less than 0.5 log unit enhancement of their sensitivity relative to ours (Devries-Rose implies a square-root effect of luminance on contrast thresholds, or a proportionality between luminance and contrast energy thresholds). The sum of these factors predicts that their optimum should be between 0.65 to 0.15 log units below ours, consistent with what is observed[3].

In every case, the curve rises from the lowest sizes, reaching a rather broad optimum somewhere between 0.5 and 2.5 cycles. In Fig. 15 we bring together several possible estimates of receptive field width and height. The model estimates derived by Anderson and Burr from the data in Fig. 14 are shown by the solid curve. The efficiency optima we extracted from Fig. 14 are shown as points. The vertical line represents the range of possible efficiency optima obtained from the present experiments (Fig. 8). While our results (vertical line) are broadly consistent with the optima from Fig. 14 (points), they are clearly higher than the model estimates of Anderson and Burr (curve). This must be considered a substantial unresolved difference between our two studies. For emphasis, in Fig. 16 we plot the spatial frequency tuning functions estimated by Anderson and Burr at 0.1 cycle/degree for left and rightward tuned receptive fields. Since direction-selectivity derives from differential excitation of these two fields, these very broad tunings would presumably lead to a diminished sensitivity, relative to more narrowly tuned sensors.

[3]There is some question how to interpret their absolute measures, since there is vertical displacement of about 0.4 log unit between width and height in several graphs, despite the fact that they share a condition (1.5 cycles in both width and height).

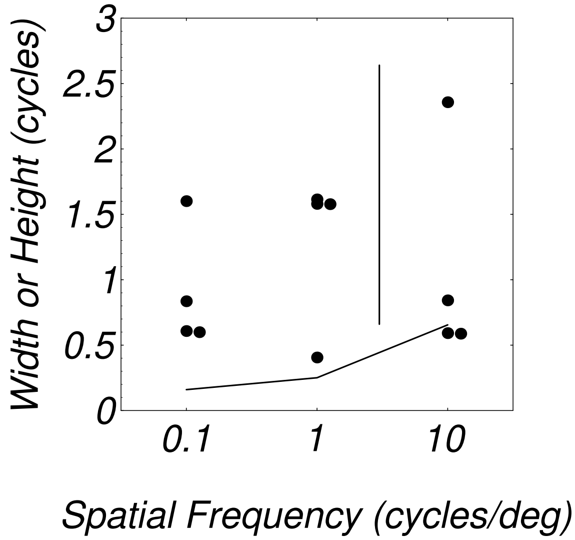

Figure 15. Comparison of receptive field size estimates from various methods. The points are efficiency optima derived from Fig. 14. The curve shows estimates obtained by Anderson and Burr from the same data from the fit of a model. The vertical line represents the range of optima observed here (Fig. 8).

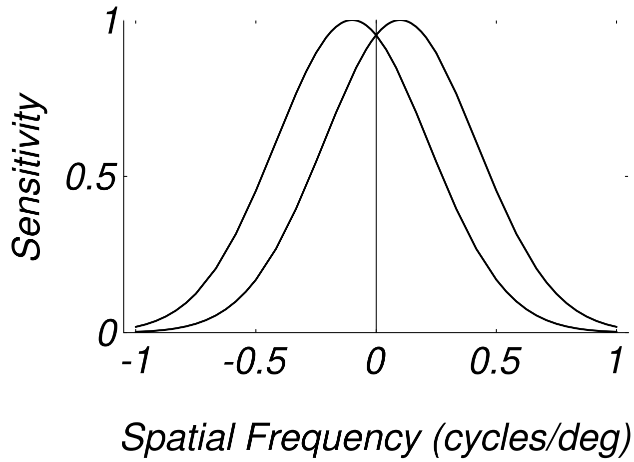

Figure 16. Spatial frequency response of left and rightward receptive fields as estimated by Anderson and Burr (1991). The peak spatial frequency is 0.1 cycle/degree.

We also note that in the optima that we derive from their data there is little systematic variation in the location of the optima with spatial frequency. This contrasts with their finding, derived by way of a detection model, that receptive fields are nearly five times as broad (in cycles) at 10 than at 0.1 cycles/deg. One possible explanation for this result is as follows. Anderson and Burr's estimates of receptive field dimensions were derived from a model incorporating sensors selective for spatial frequency, orientation, and direction. The center frequencies of the sensors were -2, -1, 0, 1, and 2 octaves relative to the test frequency. All sensors were assumed to be equally sensitive. When the stimulus is narrowed or shortened, its spatial frequency spectrum broadens, and sensors at frequencies above and below the test frequency are increasingly stimulated. If the test frequency is on a negative-sloping segment of the contrast sensitivity function, then the distribution of activity will be biased toward lower frequencies. If lower frequencies are detected by larger receptive fields, as is typically assumed, then there will be a bias toward estimation of larger receptive field dimensions. Since their model does not incorporate the variation in sensitivity with spatial frequency, it will not show this effect. The net result is that their model will overestimate receptive field dimensions at high spatial frequencies.

Subthreshold Summation

Subthreshold summation has been frequently used to probe the structure of the early human visual system (Barlow, 1958 ; Graham & Margaria, 1935; Graham & Nachmias, 1971 ; Kulikowski & King-Smith, 1973; Watson, 1982 ; Watson, et al., 1980). It relies upon the differing degrees of additivity that are associated with different types of summating mechanisms. Within a linear mechanism, linear summation is expected, while between independent detectors, probability summation is expected. In a typical experiment additivity is assessed between two signals. If summation is linear, then the signals are presumed to be detected by a mechanism that linearly sums them both. Experiments which increase the spatial or temporal extent of a signal, to discover the transition between linear and less-than linear summation, are an extension of this idea. Our experiment is perhaps the ultimate extension of this idea. Variation of the shape of the stimulus (or its spectrum) manipulates both the collection of summed components and their relative amplitudes.

However, our technique also inherits the disadvantages of subthreshold summation. Because we are attempting to measure the receptive field of one sensor that is possibly surrounded in space, spatial frequency, orientation, and temporal frequency by other sensors, the optimum is not as sharp as would be the case if this were indeed the only sensor.

Comparison with results of Watson, Barlow & Robson (1983)

In their original search for "what does the eye see best," Watson, Barlow and Robson (Watson, et al., 1983) used stimuli and methods very similar to those used here, except that a simple detection rather than a direction identification task was used. Their optimum occurred at 8 cycle/deg, 4 Hz, 2.66 cycles , and 0.142 sec, compared to our values of 3 cycle/deg, 5 Hz, 1.32 cycles, and 0.133 sec. These numbers are all quite similar , except perhaps those for spatial frequency and bandwidth. We have mentioned above that the differing frequency optima may reflect a genuine difference between the motion system and a more general detection system. The difference in spatial scale must be tempered by our observation that this parameter shows a particularly flat optimum. Their best threshold was -6.03 log deg^2 sec. This improvement over our best value (about 5.0 log deg^2 sec) may be attributed in part to an increased mean luminance (340 cd/m^2) which might yield 0.5 log unit (see discussion above regarding data of Anderson and Burr), and binocular viewing, which might yield another 0.3 log unit (Arditi, 1986; Campbell & Green, 1965).

Comparison with Cortical Receptive Fields

The motion filter model (Watson & Ahumada, 1983; Watson & Ahumada, 1985) was inspired in part by direction-selective simple cells in the visual cortex of cat and monkey (Campbell, Cleland, Cooper & Enroth-Cugell, 1968; De Valois, Yund & Hepler, 1982). More recent measurements have shown that the linear receptive fields of these simple cells are described well by the motion filter model (Hamilton, et al., 1989; McLean & Palmer, 1994; McLean, Raab & Palmer, 1994) , though their detailed behavior may require additional, non-linear mechanisms (Albrecht & Geisler, 1991; Heeger, 1994; Reid, Soodak & Shapley, 1991).

The particular form of motion filter proposed by Watson and Ahumada (Watson & Ahumada, 1983) was of a type they described as a "quadrature model," created by combining a pair of separable spatiotemporal filters in quadrature phase. Such a filter, though inseparable in space-time and frequency, is separable in frequency when only positive temporal frequencies are considered ("positive-separable"). This in turn means that the spectral receptive field would be aligned with the Cartesian axes, which likewise means that the aperture would be stationary. Hamilton et al., examining both amplitude and phase data, find general agreement with the quadrature model, and in particular with spatiotemporal separability of the spectral receptive field in one quadrant (Hamilton, et al., 1989).

McLean, Raab, and Palmer have made both space-time (McLean, et al., 1994) and frequency domain (McLean & Palmer, 1994) measurements of the receptive fields of simple cells in cat visual cortex. In agreement with Hamilton *et al.*, they found that 29 out of 30 cells showed a spectral receptive field aligned with the Cartesian axes, rather than aligned with the velocity axis.

Our optimal stimulus has a frequency bandwidth of 1.1 octaves and an orientation bandwidth of 41 degrees. These agree closely with comparable median estimates for primate cortical cells of 1.4 octaves and 42 degrees, respectively (De Valois, et al., 1982; De Valois, et al., 1982), though it must be borne in mind that the distributions of these estimates over the population of cells was very broad.

Conclusions

We measured contrast energy thresholds for a wide range of generalized Gabor stimuli, varying in spatial frequency, duration, height, width, carrier speed, and aperture speed. The lowest contrast energy threshold occurs at around 3 cycles/degree and 5 Hz, with a spatial bandwidth of about 1.1 octaves, an orientation bandwidth of about 41 degrees, and a temporal bandwidth of about 7 Hz (2.5 octaves). As an estimate of the underlying motion sensor, these values agree well with median estimates from single cortical neurons, but disagree with some other psychophysical estimates, particularly in regard to spatial bandwidth (Anderson & Burr, 1991).

We find no evidence for spectral receptive fields aligned with the velocity axis. Furthermore we point out that such a receptive field is generally incompatible with a commonplace observation: that spatial bandwidths are typically much narrower, in octaves, than temporal bandwidths.

Notation

$\mathbf{u} = [u, v, w]$	spatiotemporal frequency coordinate
$\mathbf{x} = [x, y, t]$	space-time coordinate
$\mathbf{M}$	motion matrix
$\mathbf{S}$	scaling matrix
$\mathbf{T}$	transformation matrix combining motion and scaling
$\mathbf{f} = [f_x, f_y, f_t]$	grating spatiotemporal frequency
f_t/f_x	horizontal carrier speed
s_x	horizontal aperture scale
s_y	vertical aperture scale
s_t	temporal aperture scale
a_x	horizontal aperture speed
a_y	vertical aperture speed

Acknowledgments

We thank the members of the NASA Ames Vision Group for advice and criticism. We also thank Stephen J. Anderson for comments on an earlier draft. This work supported by NASA RTOP 505-64-53 and AFOSR 91-0154.

References

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A 2*(2), 284-299.

Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature 300*, 523-525.

Albrecht, D. G. & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience 7*, 531-546.

Anderson, S. J. & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Res. 25*, 1147-1154.

Anderson, S. J. & Burr, D. C. (1987). Receptive field size of human motion detection units. *Vision Research 27*, 621-635.

Anderson, S. J. & Burr, D. C. (1989). Receptive field properties of human motion detection units inferred from spatial frequency masking. *Vision Research 29*, 1343-1358.

Anderson, S. J. & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America 8*(8), 1330-1339.

Anderson, S. J., Burr, D. C. & Morrone, M. C. (1991). Two-dimensional spatial and spatial frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America 8*(8), 1340-1351.

Arditi, A. (1986). Binocular vision. In K. Boff, L. Kaufman, & J. Thomas (Ed.), Handbook of Perception and Human Performance New York: Wiley.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology 141* , 337-350 .

Barlow, H. B. & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology (London) 178*, 477-504.

Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London, Ser. B. 227*, 249-265.

Campbell, F. W., Cleland, B. G., Cooper, G. F. & Enroth-Cugell, C. (1968). The angular selectivity of visual cortical cells to moving gratings. *Journal of Physiology 198*, 237-250.

Campbell, F. W. & Green, D. G. (1965). Monocular versus binocular visual acuity. *Nature 208*, 191-192.

Crick, F. H. C., Marr, D. C. & Poggio, T. (1981). An information processing approach to understanding the visual cortex. In F. O. Schmitt, F. G. Worden, G. Adelman, & S. G. Dennis (Ed.), The organization of the cerebral cortex Cambridge: MIT Press.

De Valois, R. L., Albrecht, D. G. & Thorell, L. G. (1982). Spatial frequency selectivity of cells in Macaque visual cortex. *Vision Research 22*, 545-559.

De Valois, R. L., Yund, E. W. & Hepler, H. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research 22*, 531-544.

Duda, R. O. & Hart, P. E. (1973). *Pattern classification and scene analysis* . New York: John Wiley.

Fleet, D. J. & Langley, K. (1993). Computational analysis of non-Fourier motion. *Vision Research in press*,

Graham, C. H. & Margaria, R. (1935). Area and intensity-time relation in the peripheral retina. *Am. Journal of Physiology 113* , 299-305 .

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. *Vision Research 11*, 251-259 .

Green, D. M. & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (1 ed.). New York: Wiley.

Hamilton, D. B., Albrecht, D. G. & Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Research 29*(10), 1285-1308.

Heeger, D. J. (1994). Modeling simple cell direction selectivity with normalized half-squared linear operators. *Journal of Neurophysiology in press*,

Kulikowski, J. J. & King-Smith, P. E. (1973). Spatial arrangement of line, edge, and grating detectors revealed by subthreshold summation. *Vision Research 13* , 1455-1478 .

Kulikowski, J. J. & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient mechanisms in human vision. *Journal of Physiology, Lond. 232* , 149-163 .

McLean, J. & Palmer, L. A. (1994). Organization of simple cell responses in the three-dimensional (3-D) frequency domain. *Visual Neuroscience 11*, 295-306.

McLean, J., Raab, S. & Palmer, L. A. (1994). Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neuroscience 11*, 271-294.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), Sensory communication New York: Wiley.

Reichardt, W. (1986). Processing of optical information by the visual system of the fly. *Vision Research 26*(1), 113-126.

Reid, R. C., Soodak, R. E. & Shapley, R. M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology 66*(2), 505-529.

Tolhurst, D. J. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *Journal of Physiology 231*, 385-402.

Tolhurst, D. J. (1975). Sustained and transient channels in human vision. *Vision Research 15*, 1151-1155.

van Nes, F. L. & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America 57*, 401-406.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A 2*(2), 300-321.

Watson, A. B. (1979). Probability summation over time. *Vision Research 19*, 515-522.

Watson, A. B. (1982). Summation of grating patches indicates many types of detector at one retinal location. *Vision Research 22*, 17-25 .

Watson, A. B. (1990). Optimal displacement in apparent motion and quadrature models of motion sensing. *Vision Research 30*(9), 1389-1393.

Watson, A. B. & Ahumada, A. J., Jr. (1983). A look at motion in the frequency domain. In J. K. Tsotsos (Ed.), Motion: Perception and representation (pp. 1-10). New York: Association for Computing Machinery.

Watson, A. B. & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A 2*(2), 322-342.

Watson, A. B., Ahumada, A. J., Jr. & Farrell, J. (1986). Window of visibility: psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A 3*(3), 300-307.

Watson, A. B., Barlow, H. B. & Robson, J. G. (1983). What does the eye see best? *Nature 302*(5907), 419-422.

Watson, A. B. & Eckert, M. P. (1994). Motion-contrast sensitivity: visibility of motion gradients of various spatial frequencies. *Journal of the Optical Society of America A 11*(2), 496-505.

Watson, A. B. & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics 33*(2), 113-120.

Watson, A. B. & Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Research 21*, 1115-1122.

Watson, A. B., Thompson, P. G., Murphy, B. J. & Nachmias, J. (1980). Summation and discrimination of gratings moving in opposite directions. *Vision Research 20*, 341-347.